

TACAM: The Threat Actor Capability & Motivation Matrix for Attributing OT Adversaries to Equipment via CWE/CPE Clusters

AUTHORS

Jim McKenney

Principal Investigator, Eigenia B.V.

PUBLISHED

Feb 13, 2026

ABSTRACT

Operational technology (OT) teams can list the vulnerabilities in their field devices, but they cannot easily say which threat actors are likely to exploit which equipment in their facility. TACAM, the Threat Actor Capability & Motivation Matrix, addresses this gap. It applies a theory developed by Jim McKenney and is in use in the Cyber Digital Twin model. TACAM profiles each actor across nine dimensions: techniques, sectors, geography, OT protocols, temporal activity, products (CPE), weaknesses (CWE), CVE and KEV exposure, and EPSS exploit pressure. It joins actor technique signatures from MITRE ATT&CK to CWE/CAPEC weakness clusters and on to National Vulnerability Database (NVD) product identifiers, and it projects each CVE to the ATT&CK techniques its weaknesses can realize. A curated cyber-incident feed derived from GDELT supplies observed attribution, which the method keeps separate from capability inferred from the weakness taxonomy. Every link carries its provenance class, and a dominance rule guarantees that inferred evidence never ranks above observed evidence. Each actor receives an Estimated Intent-Capability (EIC) score for a given facility, built from intent, capability and opportunity, scaled by CISA Known Exploited Vulnerabilities (KEV) and FIRST Exploit Prediction Scoring System (EPSS) signals, and adjusted by bounded modifiers for the remaining dimensions. The EIC score is the first factor of the twelve-factor Adversary Threat Quotient (ATQ). The current snapshot tracks 624 actors across 2,823 actor × sector affinity rows and 66,390 cyber-incident records. On a representative OT facility inventory of 6,752 CVEs, the projection resolves 613 CVEs to at least one ATT&CK technique and 372 to at least one actor group, against 5 CVEs with any technique before. Evaluation of ranking accuracy is PENDING EVALUATION.

tacam

ot-security

threat-attribution

cwe

cpe

capec

iec62443

mitre-attack

epss

kev

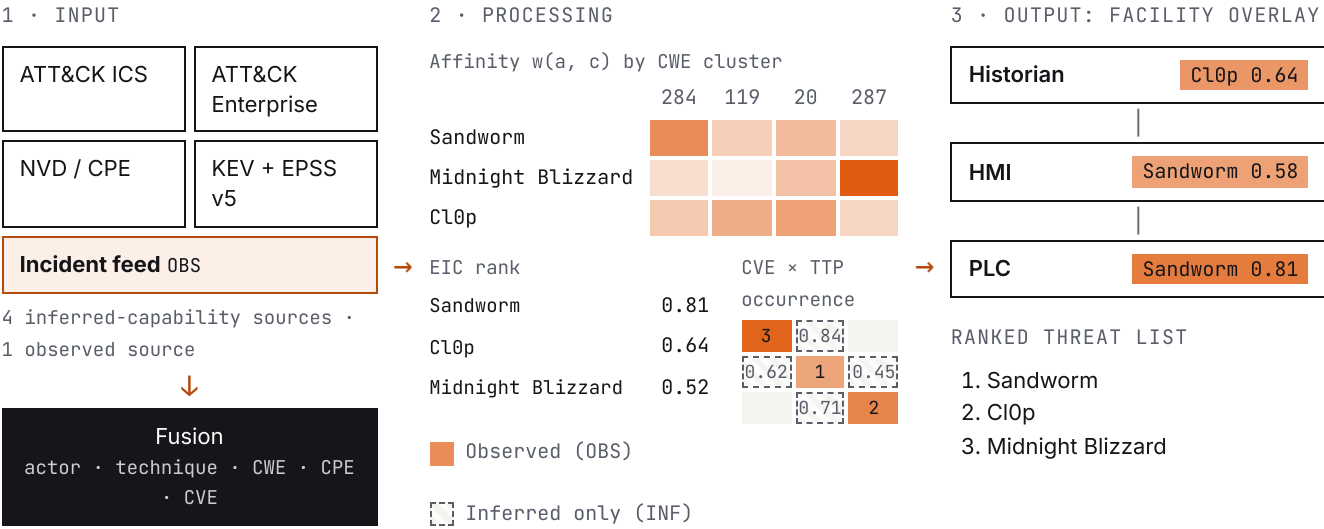
gdelt

provenance

atq

GRAPHICAL ABSTRACT

ILLUSTRATIVE VALUES



Actor technique signatures are joined through CWE clusters to products and projected CVEs, scored with KEV/EPSS and the nine-dimension profile, and overlaid on a facility inventory. Observed incident evidence is kept distinct from inferred capability throughout.

1 Introduction

Vulnerability management in OT works differently from IT. An IT team scanning a Windows fleet gets CVE scores, patch availability and, more and more often, EPSS likelihoods. An OT team scanning a distributed control system (DCS) gets the same CVSS scores applied to PLCs, safety instrumented systems and field I/O modules. Those devices cannot be patched every quarter, may run unsupported firmware for years, and sit in network segments where a failed patch can stop production or trigger a safety event.

Practitioners want to know whether any threat actor has the capability and motivation to exploit a specific device in a facility like theirs. A CVSS score answers a related but different question. A CVSS 9.8 buffer overflow in a niche SCADA historian from a vendor with three percent market share may carry less real risk than a CVSS 7.2 authentication bypass in a controller that Sandworm has shown operational interest in.

Existing frameworks each cover part of the problem. MITRE ATT&CK for ICS (Alexander et al., 2020) documents adversary techniques at the tactic level, and ATT&CK Enterprise documents the intrusion tradecraft that comes before OT access. CWE classifies weaknesses. CAPEC maps attack patterns to weaknesses and, through its taxonomy

mappings, to ATT&CK techniques. NVD links CVEs to CWEs and to CPE product strings. CISA KEV lists CVEs with confirmed exploitation, and FIRST EPSS gives each CVE a probability of exploitation. No existing method follows the whole chain from actor to technique, weakness type, product, vulnerability and facility asset to produce a scored ranking of actors for a given asset inventory.

There is also a gap in the evidence. Following that chain makes a capability argument: an actor's technique repertoire is compatible with the weakness classes in a facility's equipment. It does not show that the actor has been observed operating against those CVEs or in that sector. Treating a plausible link as an observed one inflates confidence, and OT risk assessment cannot afford that. TACAM therefore keeps two provenance classes apart throughout: capability inferred from the weakness taxonomy (INF) and attribution observed in a curated cyber-incident feed (OBS). Observed evidence can corroborate and re-weight the inferred structure, but the two are never merged into one undifferentiated score.

TACAM builds weakness clusters from CWE/CAPEC membership, scores each actor's technique use against them, joins the clusters to NVD product identifiers, projects each CVE to ATT&CK techniques, and fuses observed incident attribution into a CVE × technique × sector occurrence matrix in which every cell records its provenance. Each actor is profiled across nine dimensions, and those dimensions combine into an EIC score for a specific facility. TACAM applies a theory developed by Jim McKenney and runs as part of the Cyber Digital Twin model. This paper presents the method, the corpus as built and the descriptive results from the current snapshot. Comparative evaluation is PENDING EVALUATION.

2 Related Work

2.1 Threat Intelligence Frameworks

Strom et al. (2018) describe MITRE ATT&CK as a knowledge base of adversary behavior built from observed intrusions. Its basic unit, the technique, describes what an adversary does in enough detail to guide defenders and with enough stability to compare incidents. ATT&CK group profiles link named actors to sets of techniques. TACAM uses these profiles as its main source of actor behavior signatures.

Alexander, Belisle and Steele (2020) extended ATT&CK to industrial control systems. The ICS matrix documents tactics from Initial Access to Impact and maps them to ICS-specific techniques such as Modify Parameter, Alarm Suppression and Loss of Safety. It covers OT

intrusion phases that the Enterprise matrix did not, such as lateral movement from IT into OT networks and ICS-specific impact. TACAM uses both matrices. Enterprise techniques serve as actor signatures for the intrusion phase, and ICS techniques describe the operational phase that bears most directly on OT equipment. The tactic count changes between ATT&CK releases, so TACAM pins the release it normalizes against (§3.12).

2.2 Exploit Prediction and Prioritization

Jacobs et al. (2021) introduced EPSS as a data-driven alternative to CVSS for predicting exploitation. CVSS describes a vulnerability's severity (attack vector, complexity, privileges required) and was not designed to predict whether it will be exploited. EPSS publishes a daily probability of exploitation for each CVE from a model trained on observed exploitation events. The score distribution changes between EPSS model versions, so TACAM pins the version it uses (§3.12). The CISA KEV catalog adds a binary signal: a CVE listed in KEV has confirmed exploitation in the wild.

2.3 Attribution and Actor Profiling

Most published attribution work in OT relies on manual analysis of incident reports, malware reverse engineering and infrastructure overlap. Little academic work correlates technique sets with actor groups across large CVE corpora. TACAM scores every ATT&CK-mapped actor at once across the full corpus, where single-incident analysis examines one intrusion after the fact.

The CAPEC catalog (MITRE, 2017) maps each attack pattern to the CWE weaknesses it exploits and, in its taxonomy mappings, to the ATT&CK techniques that realize it. TACAM uses this bridge to project a CVE's weaknesses forward to techniques. As far as the author knows, no published work has used CWE clusters as the link between actor technique signatures and CPE product identifiers, which is the join at the center of TACAM.

2.4 Event Datasets as Observed Evidence

Event datasets built from open-source reporting complement the structural knowledge bases above. The GDELT Project (Leetaru & Schrod, 2013) pioneered machine-coding of global news into structured event records at scale. TACAM uses a cyber-incident derivative of this approach, in which each reported incident is structured by a language-model extraction pass into victim, country, sector, attributed actor, attributing party, affected CVEs, incident date and a flag for whether an attack was confirmed. The model and version used for each

extraction are recorded. TACAM treats this feed strictly as observed evidence: attribution asserted by named reporting. It validates the feed before use (§3.6) and keeps it in its own evidentiary register.

3 Methods

3.1 Data Sources

TACAM draws on five public sources. Four describe inferred capability, the general structure of actor tradecraft and vulnerability weakness. The fifth records observed attribution from specific reported incidents.

MITRE ATT&CK (Enterprise + ICS). Each ATT&CK group becomes an actor record, and each recorded use of a technique by a group becomes an actor × technique link. The ICS matrix is processed separately and merged on a shared set of actor names.

CWE/CAPEC taxonomy. CWE entries form a parent-child tree. CAPEC attack patterns link to CWE weaknesses through their related-weakness records and to ATT&CK techniques through their taxonomy mappings. TACAM groups CWE entries under top-level categories, for example CWE-284 (Improper Access Control) and CWE-119 (Improper Restriction of Operations within the Bounds of a Memory Buffer), and attaches each cluster's CAPEC patterns and, through them, the techniques those patterns realize.

NVD CVE/CPE feed. Each CVE record carries a primary CWE, a full weakness list, CPE product strings and CVSS, EPSS and KEV fields. The CWE places the CVE in a weakness cluster and, through §3.5, links it to techniques. The CPE string names the affected product by vendor, product name and version range.

CISA KEV and FIRST EPSS. KEV supplies a daily list of CVEs with confirmed exploitation. EPSS supplies a daily probability per CVE, and the daily history gives each CVE an EPSS trajectory.

Cyber-incident feed (observed). Each record is one reported incident with its victim, country, sector, critical-infrastructure flag, effects, intensity, CVEs, actor, attributing party, incident date and confirmation flag, together with the extraction model, version and timestamp and whether the source article was read. Each record is a claim about an observed event, not a general statement of capability.

3.2 Actor Alias Resolution NEW

The incident feed names over 1,500 distinct actors, while ATT&CK profiles 624 groups. Reporting uses many names for one group (for example APT28, Fancy Bear and Forest Blizzard) and also vague labels such as "Russian hackers". A resolution function maps each feed name n to an ATT&CK group:

$$\begin{aligned} \text{resolve}(n) &= g \text{ if } n \text{ matches an ATT\&CK alias of } g \text{ or a curated alias entry for } g, \\ &\text{else } \perp \end{aligned} \tag{1}$$

Each resolved record stores whether the match was exact or curated. Names that resolve to $\perp$ remain in the observed register and can appear in occurrence-matrix cells, but they do not contribute to any actor's EIC score.

3.3 Actor $\times$ Weakness-Cluster Affinity Matrix

For each actor a and CWE cluster c , TACAM computes an affinity weight $w(a, c)$ in two steps. First, each technique t is mapped to one or more clusters through the CAPEC chain: technique $\rightarrow$ CAPEC attack pattern $\rightarrow$ CWE weakness $\rightarrow$ cluster. When a technique maps to several clusters, its weight is split between them in proportion. Second, the techniques attributed to actor a are aggregated. Each technique is weighted by its attributed frequency in ATT&CK group profiles where frequency data exists, and given unit weight where only presence is recorded. Each cluster's weight is the sum of the technique weights mapped to it, normalized so that each actor's weights form a distribution. The result is an actor $\times$ cluster matrix W whose rows each sum to 1.0.

3.4 Cluster-to-CPE Join

Each CWE cluster is linked to the CPE product strings of every CVE whose CWE falls in that cluster. A cluster c therefore maps to a set of products $P(c)$: the equipment in which weaknesses of that type have been observed. Actor a 's affinity for product p is the sum, over clusters c , of $w(a, c)$ multiplied by the share of that cluster's CVEs in the full NVD corpus that affect product p .

3.5 CVE-to-Technique Projection

Let V be the CVE universe and T the ATT&CK technique universe (Enterprise ∪ ICS, including sub-techniques). For a CVE v , TACAM defines the projection $M(v)$ to the techniques its weaknesses can realize:

$$CWE(v) = \{ \text{primary CWE of } v \} \cup \text{weaknesses}(v), \text{ excluding CWE-Other and NVD-CWE-noinfo} \quad (2)$$

$$\Gamma(w) = \{ p \in CAPEC : w \in \text{related weaknesses of } p \} \quad (3)$$

$$\Theta(p) = \{ t \in T : t \in \text{ATT\&CK mappings of } p \} \quad (4)$$

$$M(v) = \bigcup_{w \in CWE(v)} \bigcup_{p \in \Gamma(w)} \Theta(p) \quad (5)$$

TACAM reads the ATT&CK mappings in (4) from MITRE's CAPEC distribution itself. Technique IDs there have no T prefix (for example, 1574.010 is T1574.010), so they are normalized on ingest. The CAPEC references in ATT&CK's own data cannot replace them: only a few dozen techniques carry one, most of those IDs are deprecated, and they give almost no usable coverage.

Each projected link (v, t) carries a confidence weight that rewards techniques reachable through several independent paths:

$$\phi(v, t) = 1 - \prod_{\pi \in \text{paths}(v, t)} (1 - \rho(\pi)) \quad (6)$$

Here $\text{paths}(v, t)$ counts distinct CAPEC patterns that connect v to t , so a CWE with many CAPEC links cannot inflate ϕ by repeating the same pattern. The default per-path reliability is $\rho = 0.6$, and paths from a low-precision CWE assignment get $\rho = 0$. Because $\rho < 1$, ϕ is always strictly below 1. Every link produced here is inferred (INF): it states that the vulnerability's weakness class is compatible with the technique, not that any actor was observed using t against v .

3.6 Observed Incident Evidence

Let R be the raw incident record set. A record r is valid when all four conditions hold:

1. the source article was read and the article reports a confirmed attack;
2. the incident date is not the placeholder used when no date could be extracted;

3. each CVE in the record exists in NVD (CVEs that do not resolve are dropped from the record);
4. the record's sector maps to one of the 16 CISA critical infrastructure sectors through a vocabulary-reconciliation function *map_sector*.

The observed set is the expansion of valid records over their resolved actors and CVEs:

$$O = \{ (a, s, g, v, \tau) : r \text{ valid}, a = \text{resolve}(\text{actor}(r)), s = \text{map_sector}(\text{sector}(r)), g = \text{country}(r), v \in \text{CVEs}(r) \cap \text{NVD}, \tau = \text{date}(r) \} \quad (7)$$

Each tuple is an observed attribution (OBS). *O* has no technique coordinate. The feed records that actor *a* was reported against CVE *v* in sector *s* and country *g* at time τ , not that *a* used a particular technique. The technique link for *v* remains the inferred projection $M(v)$. TACAM therefore never asserts an observed actor-technique or CVE-technique pair. Victim organization names are used only to de-duplicate records and are not reported.

3.7 CVE × TTP × Sector Occurrence Matrix

The occurrence matrix Λ holds cells $c = (v, t, s)$ with $t \in M(v)$. The observed incident count for a CVE in a sector is

$$n_{obs}(v, s) = |\{ r \text{ valid} : v \in \text{CVEs}(r), \text{map_sector}(\text{sector}(r)) = s \}| \quad (8)$$

$$occ(c) = n_{obs}(v, s) \text{ if } n_{obs}(v, s) > 0, \text{ else } \varphi(v, t) \quad (9)$$

Every observed cell scores at least 1 and every inferred-only cell scores below 1, so in any descending sort an observed cell always ranks above an inferred one, and inferred cells are ordered among themselves by confidence. Each cell also records:

- **actor names**: the actors observed against *v* in sector *s*. Inferred-only cells have an empty set, never an actor imputed from affinity.
- **provenance ledger**: one entry for each contributing incident report, plus the cluster and CAPEC pattern behind the inferred path, so a reader can see how much of a cell's weight is observed and how much inferred.
- **tactic** of the technique, and **first seen**, **last seen** and **week** from the observed incident dates. Inferred-only cells carry the snapshot date and no first or last sighting.

The matrix is rebuilt additively at each snapshot. Extending the projection from the historical CVE set to every NVD CVE with a usable weakness raises the share of a facility's CVE inventory that resolves to at least one technique, and therefore to at least one actor group.

3.8 The Nine-Dimension Actor Profile

Each actor a has a profile vector $A_d(a, \cdot)$ for each dimension d . Every entry carries its provenance class. Four dimensions form the core EIC terms. The other five enter as bounded modifiers (§3.9).

#	DIMENSION	SOURCE	PROVENANCE	ENTERS EIC AS
1	TTP	ATT&CK group profiles	INF	Capability C
2	Sector	Profile prior + observed set O	INF + OBS	Intent I
3	CWE	Affinity matrix W	INF	Opportunity O
4	CPE	Cluster-to-CPE join	INF	Opportunity O
5	Geography NEW	Victim country in O	OBS	Modifier m_{geo}
6	Temporal NEW	Incident dates in O	OBS	Modifier m_{tmp}
7	Protocol NEW	ICS techniques + protocol tags on CVE/CPE	INF	Modifier m_{pro}
8	CVE/KEV exposure NEW	Projection M + KEV	INF + OBS	Modifier m_{cve}
9	EPSS pressure NEW	EPSS trajectory	INF	Modifier m_{eps}

3.9 Estimated Intent-Capability (EIC) Scoring

A facility F is described by its sector s_F , country g_F , CPE inventory, CVE set V_F and protocol set. The core terms follow the earlier method.

Intent (I) blends a profile prior with observed incident frequency, where $\hat{I}_{obs}(a, s)$ is the actor's resolved incident count in sector s divided by its largest count in any sector:

$$I(a, s) = \lambda \cdot \hat{I}_{profile}(a, s) + (1 - \lambda) \cdot \hat{I}_{obs}(a, s)$$

(10)

Capability (C) scores technique breadth and sophistication from the actor's ATT&CK technique count, weighted by tactical diversity and by the presence of ICS techniques.

Opportunity (O) is the actor $\times$ cluster affinity $w(a, c)$ joined to the CPE products in F .

Modifiers. Each of the five modifier dimensions has a facility score $\hat{s}_d(a, F)$ in $[0, 1]$, and the modifier is

$$m_d(a, F) = 1 + \kappa_d \cdot (2 \cdot \hat{s}_d(a, F) - 1) \text{ if evidence exists, else } 1 \quad (11)$$

A modifier equals 1 when there is no evidence, so an actor is never penalized for gaps in a data source. With the default $\kappa_d = 0.10$, each modifier stays within $[0.90, 1.10]$. The five scores are:

- **Geography:** $\hat{s}_{\text{geo}} = \hat{G}(a, g_F)$, the actor's resolved incident share in country g_F , normalized so that its top country is 1.0.
- **Temporal:** $\hat{s}_{\text{tmp}} = R(a) = 2^{-\Delta t(a)/h}$, where $\Delta t(a)$ is the number of days since the actor's last resolved incident and $h = 180$ days. A dormant actor is reduced by at most 10% and is never removed.
- **Protocol:** $\hat{s}_{\text{pro}} = \Psi(a, F) = |\text{protocols}(a) \cap \text{protocols}(F)| / |\text{protocols}(F)|$.
- **CVE/KEV exposure:** $\hat{s}_{\text{CVE}} = \sum_{v \in V_{a,F}} \omega(v) / \sum_{v \in V_F} \omega(v)$, where $V_{a,F}$ is the set of facility CVEs whose projected techniques $M(v)$ overlap the actor's techniques, and $\omega(v) = 2$ for KEV-listed CVEs and 1 otherwise.
- **EPSS pressure:** $\hat{s}_{\text{eps}} =$ the percentile rank, among actors scored for F , of the mean 30-day change in EPSS score over $V_{a,F}$. An actor at the median gets a modifier of exactly 1.

Exploitation factor. As before, the score is also multiplied by $X(a, F)$, which blends the share of KEV-listed CVEs in the actor's CPE intersection with the mean EPSS score of that intersection and ranges from 1.0 to a ceiling of 2.0. The full score is

$$EIC(a \mid F) = [\alpha \cdot I(a, s_F) + \beta \cdot C(a) + \gamma \cdot O(a, F)] \cdot X(a, F) \cdot \prod_d m_d(a, F) \quad (12)$$

Overlap. KEV and EPSS reach the score twice: KEV through X and m_{CVE} , and EPSS through the level in X and the trajectory in m_{eps} . Both routes are kept deliberately and are bounded. Together the five modifiers can move a score by a factor between 0.59 and 1.61, and X by up to 2.0. Whether the second route adds information beyond the first is one of the ablations in §5.4.

3.10 Reported With Every Ranked Actor

Four diagnostics are reported with each ranked actor so that a reader can see why the actor is on the list. They do not change the score.

$$K(a, F) = \left| \text{tactics}(\text{techniques}(a) \cap M(V_F)) \right| / 15$$

(13)

$$E_{obs}(a, F) = \text{observed share of the score mass in (12)}$$

(14)

$$\Delta \hat{I}_{obs}(a, s) = \hat{I}_{obs}^{[k]}(a, s) - \hat{I}_{obs}^{[k-1]}(a, s)$$

(15)

Kill-chain coverage K counts the Enterprise tactics an actor can reach through the facility's CVEs, out of the 15 in the pinned ATT&CK release. The evidence share E_{obs} shows how much of a rank rests on observation. The trend $\Delta \hat{I}_{obs}$ compares observed sector intensity in the current window $[k]$ with the previous window $[k-1]$; it defines the campaign-momentum score and drives the convergence query in §5.2. The EIC term breakdown is reported alongside these.

3.11 Parameters

All defaults below are set by judgment and are pending calibration against the evaluation in §5.4. The modifier weights κ_d are equal by design until that evaluation supports different values.

SYMBOL	DEFAULT	MEANING
α, β, γ	0.35, 0.35, 0.30	Weights of intent, capability and opportunity
λ	0.5	Profile prior vs observed intent
κ_d	0.10 each	Maximum swing of each of the five modifiers
h	180 days	Recency half-life
ρ	0.6	Reliability of each projection path (distinct CAPEC patterns)
$X_{\max}$	2.0	Ceiling of the KEV/EPSS exploitation factor
ω	2 / 1	Weight of KEV-listed vs other CVEs in $\hat{s}_{cve}$

3.12 Operation in the Cyber Digital Twin

TACAM runs inside the Cyber Digital Twin model, which supplies the facility description F . KEV and EPSS inputs refresh daily, the incident feed is revalidated and the occurrence matrix rebuilt at each snapshot, and actor technique data updates with each ATT&CK release. Two external versions are pinned because they change scores: the current ATT&CK Enterprise release, which lists 15 tactics and sets the denominator of K , and EPSS model v5, whose score distribution sets the percentile ranks in $\hat{s}_{\text{eps}}$ and the levels in X . Sector reach is normalized against the 16 CISA critical infrastructure sectors.

4 Results

4.1 Corpus Scale

As of the snapshot date (13 October 2026), the TACAM corpus contains:

COUNT	RECORD
624	tracked threat actors , each with at least one ATT&CK technique attribution.
2,823	actor × sector affinity rows : distinct actor-sector pairs with a non-zero targeting score.
624	actors with a campaign-momentum score, one per tracked actor.
66,390	cyber-incident records in the observed feed, of which 34,322 report a confirmed attack. The feed covers 14 of the 16 CISA sectors and names over 1,500 distinct actors. Only a minority of records carry CVEs that resolve in NVD, so the observed layer is dense in actor and sector attribution and sparse in CVE linkage.
17,210	occurrence-matrix cells before the projection extension, across 208 CVEs and 198 techniques, with no actor attribution. Projection coverage after the extension is reported in §4.4. The share of feed names that resolve to ATT&CK groups is PENDING EXECUTION .

Profile counts for the five new dimensions and distribution statistics across affinity deciles are PENDING EVALUATION.

4.2 High-Affinity Actor × Sector Pairings

The highest-scoring actor × sector pairs in the current snapshot include:

- **Midnight Blizzard → information-technology sector:** targeting score ≈ 0.955 , the highest single actor-sector score in the corpus. Midnight Blizzard (also tracked as NOBELIUM and Cozy Bear) has consistent ATT&CK attribution for credential access, spearphishing for initial access and supply-chain compromise, techniques that map closely to IT operator environments.
- **Sandworm/ELECTRUM → energy sector:** the strongest energy-sector affinity among ICS-active actors, in line with Sandworm's record of disruptive ICS operations, including attacks on the Ukrainian power grid.

Other actors with high campaign momentum include **CIOP**, a ransomware operator that targets a wide range of critical-infrastructure sectors, and **ALPHV/BlackCat**, a ransomware-as-a-service group with affinity for energy and manufacturing. Where both the ATT&CK prior and repeated reported incidents place an actor in a sector, the blended intent in (10) reinforces the ranking. Where only the prior does, the pairing rests on capability inference and the provenance ledger says so.

4.3 EIC Score Distribution

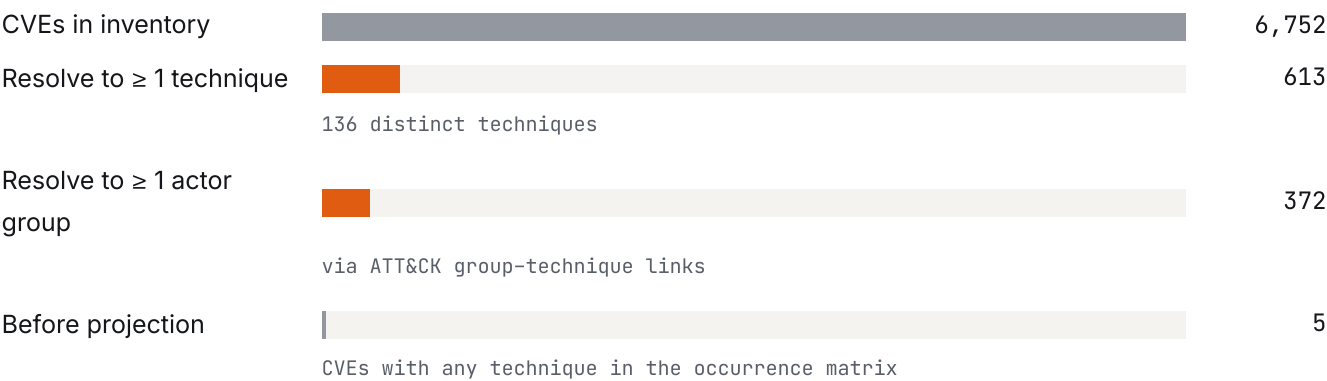
The distribution of EIC scores across the 624 actors, broken down by core term and modifier, its sensitivity to λ and κ_d , and its validation against confirmed attribution are PENDING EVALUATION.

4.4 Cluster and Projection Coverage

The projection in §3.5 has been run against the current corpus, using the CAPEC distribution as fetched on 13 October 2026. It holds 1,212 CAPEC-to-CWE related-weakness pairs across 450 CAPEC patterns and 336 CWEs, and 271 CAPEC-to-technique ATT&CK mappings across 177 CAPEC patterns and 189 techniques.

On a representative OT facility inventory of 6,752 distinct CVEs, the projection resolves 613 CVEs to at least one ATT&CK technique, covering 136 distinct techniques. Of these, 372 CVEs resolve onward to at least one attributed ATT&CK group through group-technique links. Before the projection, only 5 CVEs in the occurrence matrix carried any technique.

PROJECTION ON A REPRESENTATIVE OT FACILITY INVENTORY



§4.4 CVE-to-technique projection coverage. Bars share one scale: the full inventory.

The resolved techniques are spread across many families instead of collapsing onto a few frequent impact techniques, and they follow the weakness class of the source CVE. Resource-exhaustion weaknesses lead to endpoint denial of service (T1499.*), permission and path weaknesses lead to hijack execution flow (T1574.*), and a further group leads to exploitation for defense evasion (T1211). This is a qualitative check that the projection carries weakness-class signal and not co-occurrence noise.

Per-cluster product density, cluster-level EPSS and KEV statistics, and the ratio of observed to inferred cells once incident fusion (§3.6–§3.7) is materialized are PENDING EVALUATION.

5 Discussion

5.1 Operational Implications

Without TACAM, an OT team can ask which of its vulnerabilities are high severity. With TACAM, it can ask which threat actors matter most for its asset inventory, and which CVEs in that inventory match those actors' known technique affinity. For facilities that cannot patch on an IT schedule, the second question leads more directly to a decision.

A facility with a large installed base from a vendor in the energy-sector product namespace would see Sandworm at the top of its list because Sandworm's technique affinity, the product vulnerabilities in that inventory and the KEV/EPSS weight on those CVEs all point the same way. Sandworm's general reputation plays no part in the score. The occurrence matrix lets

the same team trace any CVE in its inventory to the techniques its weaknesses can realize and from there to the groups that use them, and the provenance ledger shows at each step whether the link is observed or inferred.

5.2 Query Patterns

Three queries cover most practitioner use. Each returns the ranked actors with their EIC term breakdown and the diagnostics of §3.10.

QUERY	DIMENSIONS	INPUT	OUTPUT
Q1 Inventory match	Sector × CPE × TTP, all modifiers	Facility F	Actors ranked by $\text{EIC}(a \mid F)$, with K , Ψ and E_{obs}
Q2 Vendor delta	CPE, CVE/KEV exposure	F and a product change F'	Actors entering or leaving the list above a threshold, change in KEV-listed CVEs, change in dominant CWE clusters
Q3 Sector convergence	Temporal × sector (observed only)	Sector s , windows $[k-1]$, $[k]$	Actors sorted by $\Delta \hat{I}_{obs}(a, s)$

5.3 Relation to the Adversary Threat Quotient

TACAM's output feeds the twelve-factor Adversary Threat Quotient (ATQ), which is the actor score consumed by the rest of the Cyber Digital Twin. EIC is ATQ's first and largest factor, and most of the other factors draw on individual TACAM dimensions. The ATQ weights are actuarial judgment, not a fitted result. The crosswalk below shows where each ATQ factor gets its input.

#	ATQ FACTOR	WEIGHT	TACAM INPUT
1	EIC Score	18%	EIC($a \mid F$), equation (12)
2	Kill Chain Completeness	14%	TTP; K over 15 tactics
3	Temporal Threat Score	8%	Temporal
4	EPSS Base	10%	EPSS (v5) level over $V_{a,F}$
5	Technique Reach	10%	TTP
6	Vendor Exposure	10%	CPE
7	Sector Reach	5%	Sector, over 16 CISA sectors
8	Protocol Reach	5%	Protocol
9	Incident Count	5%	Observed set O
10	Campaign Recency	5%	Temporal, $R(a)$
11	EPSS Velocity	5%	EPSS trajectory
12	Geopolitical Pressure	5%	Geography; ACLED conflict data (outside TACAM)

The weights sum to 100%. Because factor 1 already contains every TACAM dimension, factors 2 to 12 count several dimensions a second time. ATQ accepts this overlap as a design choice; it is stated here so that readers do not treat the factors as independent.

Beyond ATQ, the Cyber Digital Twin passes actor scores to a Monte Carlo attack-path simulation, an annualized loss expectancy (ALE) model and a Gordon-Loeb investment model. Those models are outside the scope of this paper.

5.4 Evaluation Protocol and Baselines

The observed layer could supply ground truth for evaluating the ranking, but it also feeds the intent term, so a naive evaluation would be circular. TACAM's evaluation therefore uses a temporal holdout. The model is fitted on observed incidents dated before a cutoff τ_0 . The actor ranking for each sector is then scored against validated incidents after τ_0 , counting only those whose attributing party differs from the reporting source. Metrics are precision@k and nDCG@k per sector. The baselines are a uniform ranking over high-CVSS CVEs, a profile-only ranking ($\lambda = 1$) and the score without modifiers ($\kappa_d = 0$). Dropping each modifier in turn, and each of the two KEV/EPSS routes, measures what each adds. This evaluation is PENDING EVALUATION.

5.5 Integration with IEC 62443 Risk Assessment

IEC 62443-3-2 requires a threat assessment that considers the likelihood of an attack as a function of actor motivation and capability against specific security level targets. TACAM's EIC score can feed this assessment. The ranking for a zone supplies motivation and capability, and the CPE intersection with that zone's inventory ties the assessment to the products actually present. The split between observed and inferred evidence matches the difference between likelihood supported by evidence and likelihood estimated from capability. The Cyber Digital Twin model documents this integration, but it has not yet been checked against a completed IEC 62443-3-2 assessment.

5.6 Limitations

Ranking accuracy not yet evaluated. No precision, recall or ranking-accuracy claim can be made until the protocol in §5.4 is run. The results in §4 are descriptive only.

Parameters set by judgment. Every default in §3.11 is a judgment. The equal modifier weights are a placeholder, not a finding.

Observed layer is sparse in CVE linkage. Few incident records carry resolvable CVEs, so the observed boost in (9) and the observed terms in (10) concentrate on a small set of widely reported CVEs. The long tail of the occurrence matrix stays inferred-only, and an inferred-only cell should be read as a capability hypothesis.

Extraction and alias quality. The incident feed comes from automated extraction over open-source reporting, and errors in actor, CVE or sector propagate into the observed layer. Validation, recorded extraction provenance and alias match types limit this but do not re-verify each claim. Feed names that do not resolve fall out of EIC entirely, which favors well-known groups.

Projection bias. Counting distinct CAPEC patterns limits, but does not remove, the advantage that common weaknesses such as CWE-20 have in ϕ . CAPEC's ATT&CK mappings are also incomplete, so some weaknesses do not project to any technique.

Protocol derivation. NVD has no standard protocol field. Protocol tags are derived from ICS techniques and product descriptions and are marked inferred.

Double counting. KEV and EPSS reach EIC through two routes, and ATQ counts TACAM dimensions again in factors 2 to 12. Both overlaps are bounded and stated, and both are tested by ablation.

ATT&CK completeness and lag. Group profiles reflect public reporting, so actors with little reporting are underrepresented. Technique updates follow the ATT&CK release schedule, and a change in the tactic count between releases shifts K for every actor.

CWE and CPE quality. CVEs assigned CWE-Other or NVD-CWE-noinfo drop out of the cluster join and the projection. CPE strings vary in detail between vendors; fuzzy matching reduces but does not remove the resulting gaps.

6 Conclusion

TACAM links threat actors to OT equipment through CWE/CAPEC weakness clusters, projects each CVE to the techniques its weaknesses can realize, and combines that capability structure with observed incident attribution under a strict provenance rule. Each actor is profiled across nine dimensions. Four form the core of the EIC score and five adjust it through bounded modifiers that stay neutral when evidence is missing. The EIC score is the first factor of ATQ.

The current corpus tracks 624 actors across 2,823 actor × sector affinity pairs and holds 66,390 incident records. High-affinity pairs such as Midnight Blizzard → IT (0.955) and Sandworm → energy show that the affinity layer separates actors by sector in ways that match their documented histories.

In practice, TACAM lets an OT team rank threats by the actors likely to target its specific inventory instead of by generic severity, and see for each rank how much rests on observation. The projection has now been run: on a representative facility inventory it raises the number of CVEs with at least one technique from 5 to 613. The next steps are to materialize incident fusion in the occurrence matrix, measure alias resolution and dimension coverage, and run the temporal-holdout evaluation that will calibrate the parameters.

References

- Alexander, M., Belisle, M., & Steele, J. (2020). *MITRE ATT&CK for ICS* (MITRE Technical Report). MITRE Corporation. <https://attack.mitre.org/matrices/ics/>
- CISA. (n.d.). *Known Exploited Vulnerabilities Catalog*. Cybersecurity and Infrastructure Security Agency. <https://www.cisa.gov/known-exploited-vulnerabilities-catalog>
- CISA. (2024). *Critical Infrastructure Sectors* (16 sectors under PPD-21, retained by National Security Memorandum 22, April 2024). Cybersecurity and Infrastructure Security Agency.

FIRST. (n.d.). *Exploit Prediction Scoring System (EPSS)*, model v5. Forum of Incident Response and Security Teams. <https://www.first.org/epss/>

Jacobs, J., Romanosky, S., Edwards, B., Adjerid, I., & Roytman, M. (2021). Exploit prediction scoring system (EPSS). *Digital Threats: Research and Practice*, 2(3), Article 20. <https://doi.org/10.1145/3436242>

Leetaru, K., & Schrod, P. A. (2013). GDELT: Global Data on Events, Location, and Tone, 1979–2012. *Annual Meeting of the International Studies Association*. <https://www.gdelproject.org/>

MITRE Corporation. (2017). *Common Attack Pattern Enumeration and Classification (CAPEC)*. <https://capec.mitre.org/>

MITRE Corporation. (n.d.). *MITRE ATT&CK Enterprise Matrix*, current release (15 tactics). <https://attack.mitre.org/>

Strom, B. E., Applebaum, A., Miller, D. P., Nickels, K. C., Pennington, A. G., & Thomas, C. B. (2018). *MITRE ATT&CK: Design and philosophy* (MITRE Technical Report MTR190003). MITRE Corporation. https://attack.mitre.org/docs/ATTACK_Design_and_Philosophy_March_2020.pdf

Figures

Figures marked illustrative show layout and encoding with placeholder values until the corresponding data exports are complete.

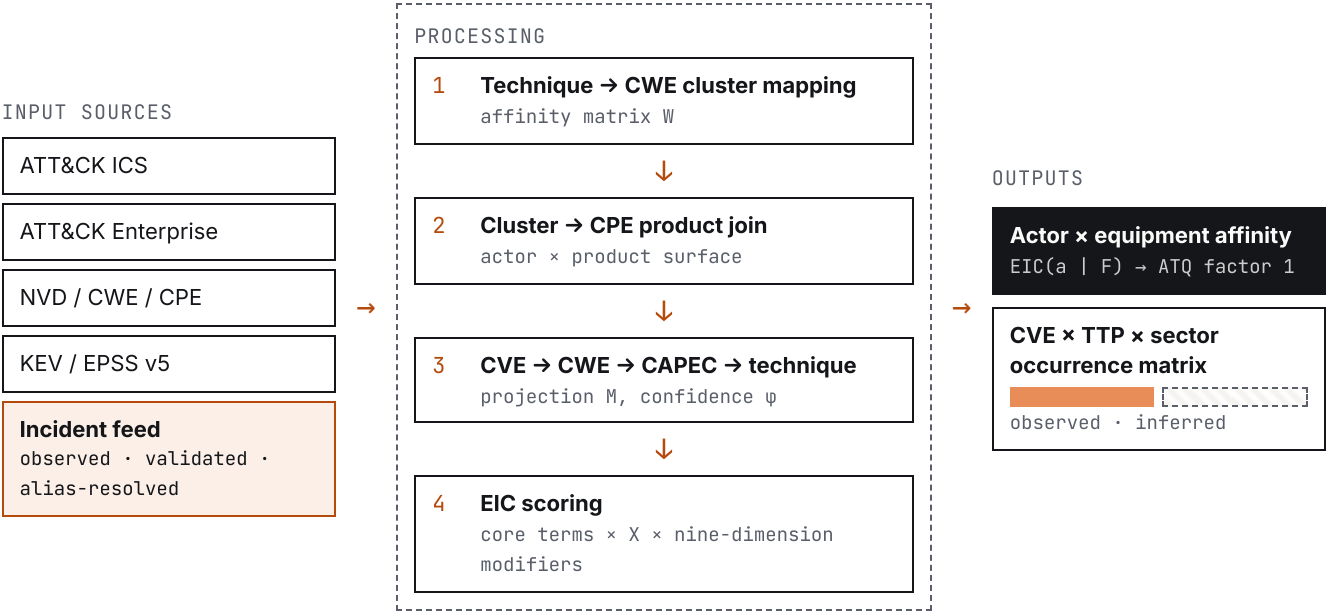


FIG. 1 TACAM Pipeline Architecture. Five sources, four of inferred capability and one of observed attribution, feed four processing stages. Outputs are the actor × equipment affinity (EIC) and the CVE × technique × sector occurrence matrix, with observed cells solid and inferred cells hatched.

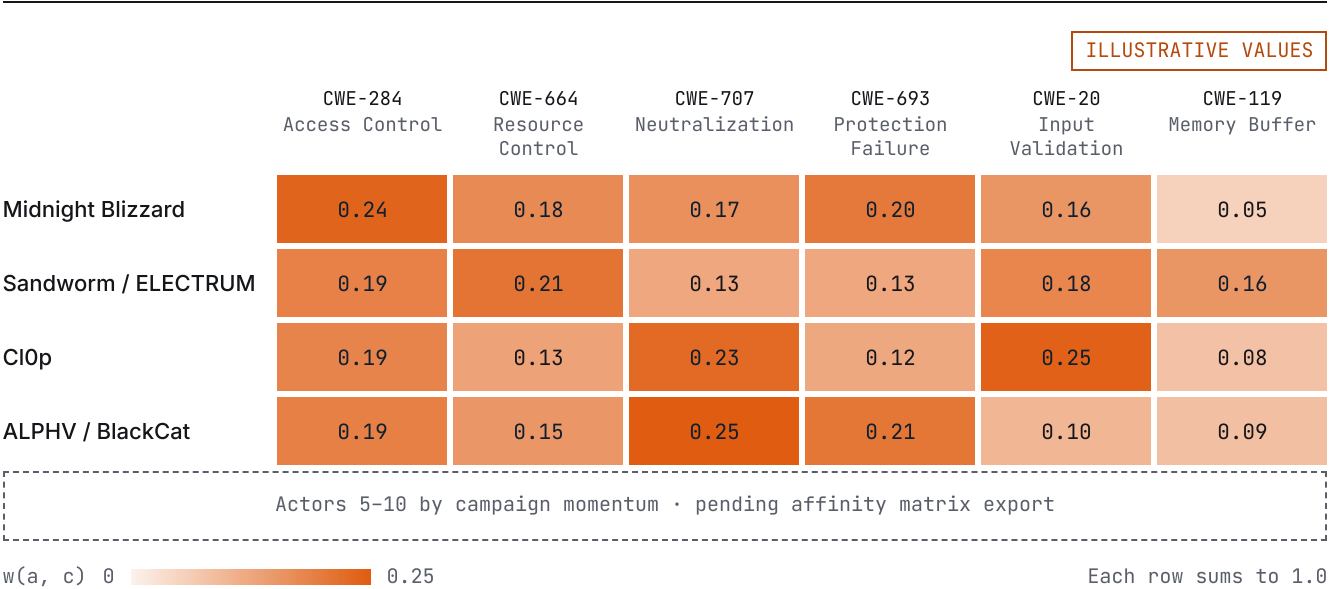


FIG. 2 Actor × CWE Cluster Affinity Heatmap. Representative actors against CWE top-level clusters. Cell intensity encodes affinity weight $w(a, c)$. Rows sort by total EIC and columns by CVE count per cluster once the full affinity matrix is exported.

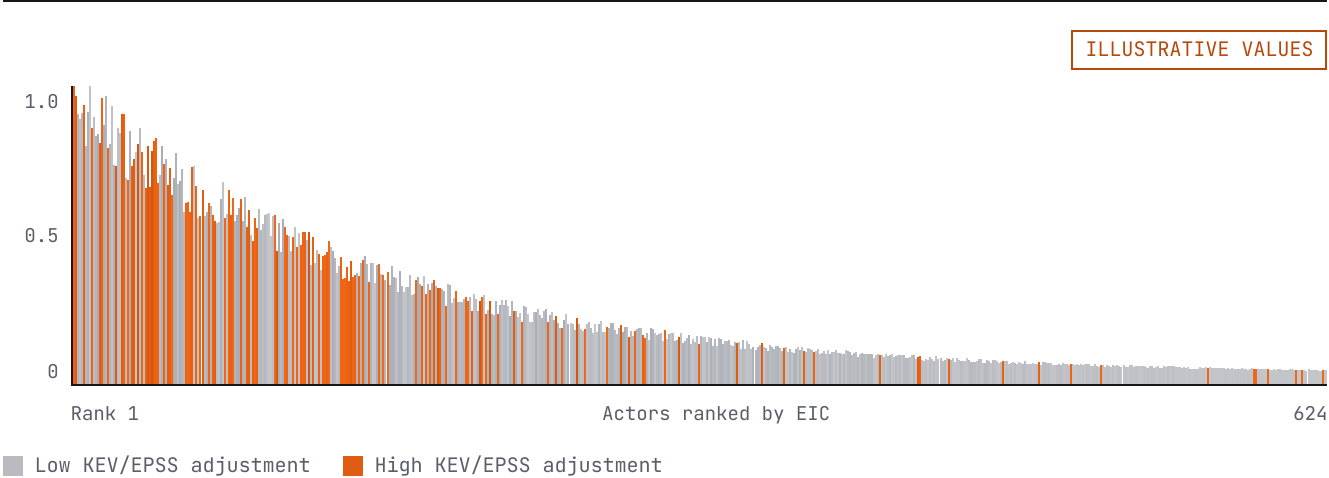


FIG. 3 EIC Score Distribution. EIC scores for all 624 actors, colored by the KEV/EPSS exploitation factor. Pending data: full EIC score export. The figure separates actors with high technique affinity but little exploitation signal from those with moderate affinity and confirmed KEV/EPSS presence.

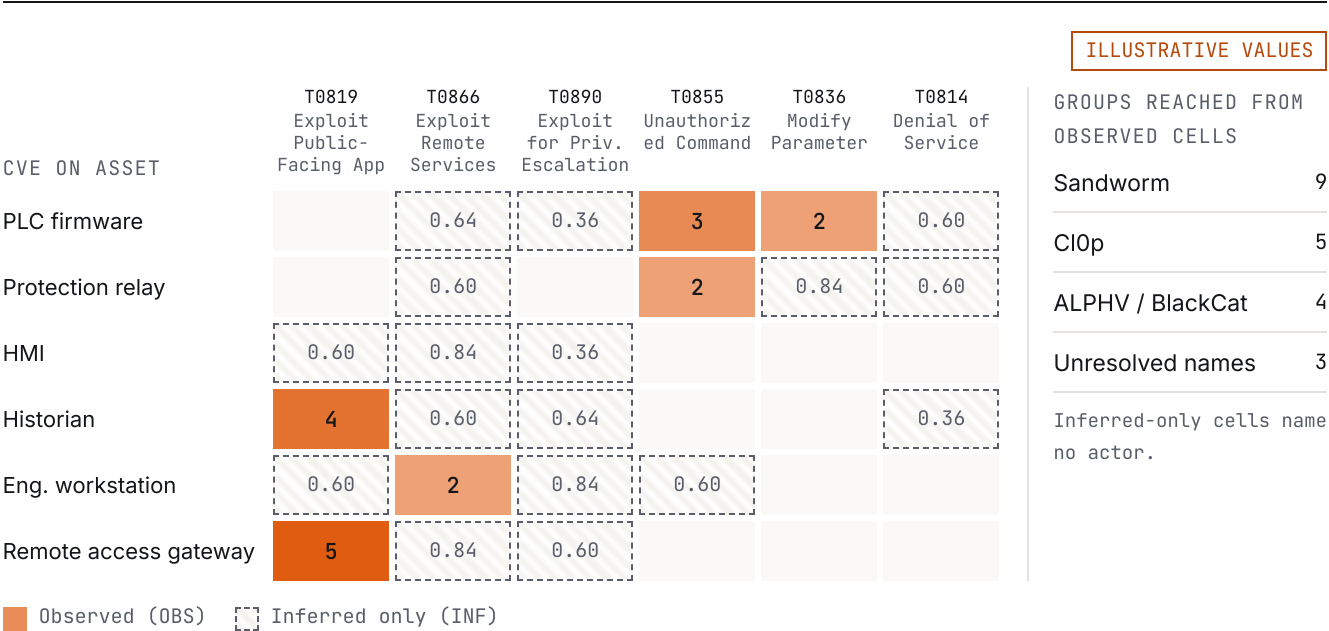


FIG. 4 Occurrence Matrix Provenance View. A CVE × technique grid for a representative facility. Observed cells show the incident count and carry named actors. Inferred-only cells show the projection confidence ϕ , always below 1, and carry no actor. Pending: representative inventory and occurrence-matrix rebuild.

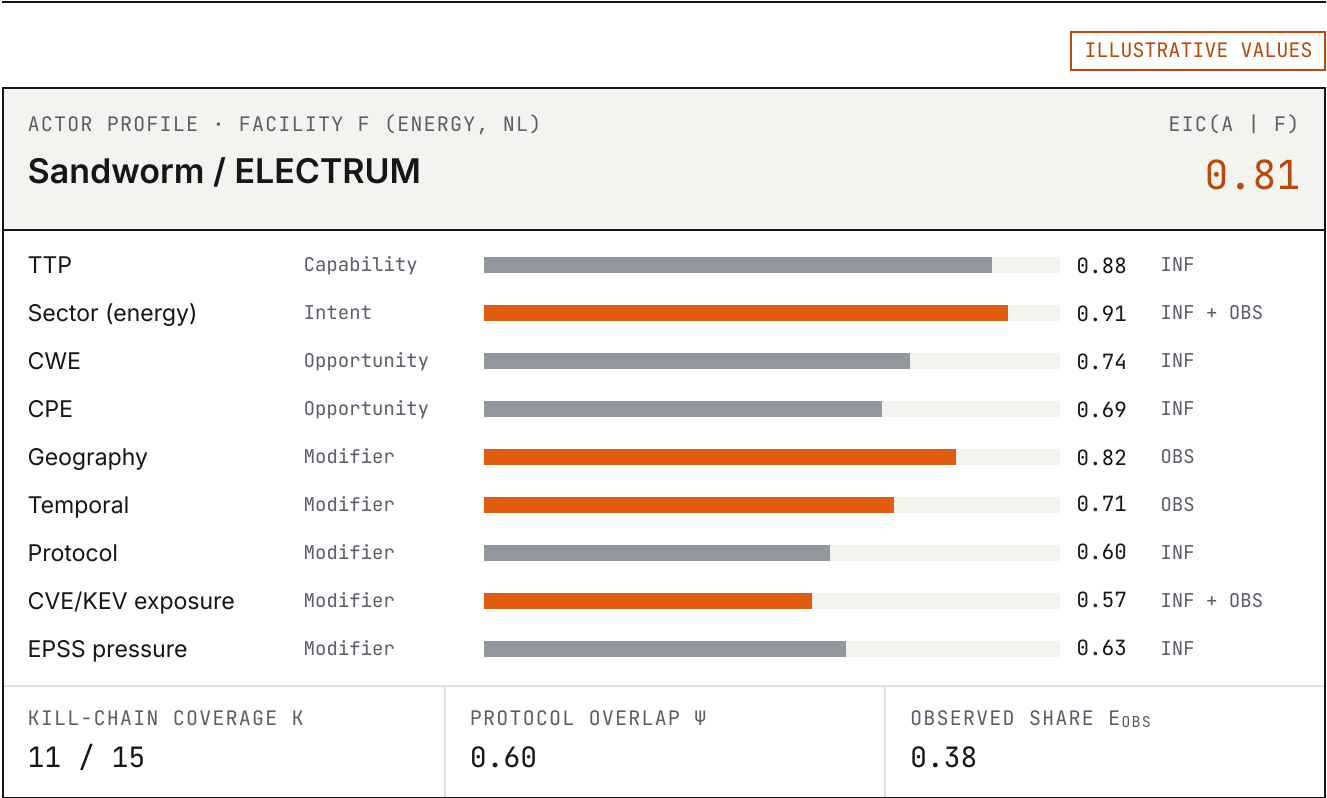


FIG. 5 Actor Profile Card (Q1 output). One ranked actor for a facility, showing each of the nine dimensions with its role in the score and its provenance class, and the diagnostics of §3.10. Bars with observed provenance are orange; inferred-only bars are gray.